

Short-Term Wind Power Prediction Method Based on Multi-feature Adaptive Fusion and Extreme Value Theory Tail Calibration

Kangxu Di

*School of Mathematical Sciences, Capital Normal University, Beijing, China
dikangxu@126.com*

Abstract. Accurate short-term wind power prediction is a critical foundation for improving the accommodation capacity and dispatching security of power systems. Aiming at the problems of multicollinearity in traditional models and large prediction errors under extreme fluctuations, this paper proposes a prediction method combining multi-feature fusion ridge regression with extreme value theory. First, multiple ridge regression branches are constructed, and multi-dimensional electrical features such as voltage and reactive power are introduced. L2 regularization is applied to alleviate variable multicollinearity, and a difference strategy is adopted to adapt to the inertial characteristics of short-term power. Second, the optimal backbone model for different prediction horizons is adaptively matched based on validation set errors, and weighted fusion with the persistence model is performed to output point prediction results. Finally, the POT-GPD model is used to fit the tail of errors, and a 95% probabilistic prediction interval is constructed. Experiments show that the average error of the proposed model across multiple horizons outperforms comparison models, and the prediction interval coverage rate exceeds 95%, which balances both conventional prediction accuracy and early warning capability for extreme wind power fluctuation risks.

Keywords: wind power prediction, ridge regression, multi-feature ridge regression, difference ridge regression, extreme value theory

1. Introduction

With the continuous expansion of new energy grid integration, the randomness, intermittency and volatility of wind power have imposed higher requirements on power system dispatching and safe operation [1, 2]. Accurate short-term wind power prediction can provide an essential basis for power grid dispatching, reserve capacity arrangement and wind power accommodation [3, 4]. However, wind power is affected by various factors such as wind conditions, unit operation status and grid-side operating conditions, and its time series usually exhibits obvious non-stationarity and abrupt change characteristics, which makes short-term prediction remain a challenging task [5, 6].

Existing wind power prediction methods can be generally categorized into statistical models, machine learning models and deep learning models. Statistical models feature simple structures and

strong interpretability, and deliver stable performance in short-term prediction tasks, such as the ARIMA model [7, 8] and multiple linear regression [9]. Machine learning methods can construct nonlinear mapping relationships using multi-dimensional operating variables, such as BP neural networks [10, 11] and random forests [12, 13]. Deep learning methods have advantages in modeling complex time series, such as LSTM [14] and convolutional neural networks [15]. Nevertheless, an easily overlooked issue in practical short-term wind power prediction is that the optimal model may vary across different prediction horizons. For very short-term prediction, the persistence model often shows strong competitiveness; as the prediction horizon increases, the persistence assumption relying solely on current power gradually fails, requiring more historical features and correction models.

In addition, conventional point prediction metrics such as MAE and RMSE only reflect the average error level, and are insufficient to characterize the risk of large errors under extreme fluctuation scenarios. Wind power prediction errors usually contain tail samples — a small number of moments with large errors exert a greater impact on power grid dispatching risks. Therefore, pursuing only the minimum average error is not enough to fully evaluate model reliability; further modeling and calibration of the prediction error tail are necessary.

To address this issue, this paper proposes a short-term wind power prediction method based on multi-feature adaptive fusion and extreme value theory tail calibration, referred to as MFBlend-EVT. This method takes multi-dimensional grid-side operating statistical variables as input, constructs candidate predictors through multi-feature ridge regression and difference ridge regression, and selects the optimal fusion structure for different prediction horizons using the validation set. On this basis, extreme value theory is applied to fit the tail of validation set residuals to form prediction intervals, thereby improving the model's coverage capability for large error risks.

2. Algorithm introduction

2.1. Ridge regression

Ridge regression [16] is a linear regression model with L2 regularization. For autoregressive modeling of wind power, the traditional least squares method cannot resolve multicollinearity and overfitting caused by multiple lag variables. The ridge regression algorithm introduces an L2 regularization term to mathematically address multicollinearity and overfitting. Ridge regression aims to minimize the following objective function:

$$J = \sum_{i=1}^m (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad (1)$$

where the first term is the sum of squared errors, the second term is the regularization term, and $\lambda \geq 0$ is the regularization coefficient. The ridge estimate of the regression coefficient obtained by optimizing the above objective function is:

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T y \quad (2)$$

where X is the design matrix and y is the target value vector. Multicollinearity in autoregressive data causes the design matrix X to be non-full rank, making $|X^T X|$ approach 0 — that is, $X^T X$ is irreversible. Ridge regression adds a regularization term λI to the $X^T X$ matrix, ensuring matrix non-singularity and thus effectively handling data multicollinearity.

Although ridge regression is a biased estimator, by selecting an appropriate regularization coefficient λ , it can achieve a balance between bias and variance and yield an estimator with smaller Mean Squared Error (MSE).

2.2. Multi-feature ridge regression

On the basis of ridge regression, the dimension of independent variables is expanded. In addition to historical active power, covariates such as voltage, current and reactive power are introduced to improve prediction accuracy by leveraging multi-dimensional data information; meanwhile, L2 regularization is retained to effectively address multicollinearity among multiple independent variables. The prediction expression is:

$$\hat{P}_{t+h} = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + \gamma_1 V_t + \gamma_2 I_t + \gamma_3 Q_t + \dots \quad (3)$$

Compared with basic ridge regression, additional features include V_t (voltage at the current moment and corresponding historical lag terms), I_t (current at the current moment and corresponding historical lag terms), and Q_t (reactive power at the current moment and corresponding historical lag terms).

Single-variable power prediction cannot distinguish different operating states with identical active power. When two moments have the same active power value but one is accompanied by rising current and falling voltage, the subsequent power change trends differ significantly. Multi-feature ridge regression accounts for the influence of these features, introduces multi-dimensional variables, captures more complete equipment status information, and improves the ability to identify power trend changes.

2.3. Difference ridge regression

For short-term prediction scenarios, to fit the inertial characteristics of short-term wind power, a difference modeling strategy is further proposed. Instead of directly predicting the absolute value of future power, difference ridge regression predicts the increment of future power relative to the current power, and the final predicted value is obtained by adding the predicted increment to the current power. The algorithm steps are as follows:

(1) Define the power variation:

$$\Delta P_{t,h} = P_{t+h} - P_t \quad (4)$$

where P_t denotes the power value at current moment t , P_{t+h} denotes the power value at future moment $t+h$, and $\Delta P_{t,h}$ denotes the power change from current moment t to future moment $t+h$.

(2) Predict the variation using multi-feature ridge regression:

$$\Delta \hat{P}_{t,h} = f(X_t) \quad (5)$$

where X_t denotes the multi-dimensional feature vector at current moment t , $f(X_t)$ denotes the ridge regression model that uses features X_t to predict the power increment, and $\Delta \hat{P}_{t,h}$ denotes the model-predicted power variation.

(3) Output the final power prediction:

$$\hat{P}_{t+h}^{\text{Delta}} = P_t + \Delta \hat{P}_{t,h} \quad (6)$$

where $\hat{P}_{t+h}^{\text{Delta}}$ denotes the final output future power prediction value.

Short-term wind power has strong autocorrelation. As mentioned for the persistence baseline model, current power already serves as a good estimate of future power. Difference ridge regression does not fit the absolute value of the predicted variable from scratch; it only needs to learn the correction amount of the current value, resulting in simpler prediction and stronger stability. In essence, difference ridge regression is a correction to the persistence baseline model: the persistence baseline model assumes the variation is always 0, while difference ridge regression allows the variation to be dynamically predicted by multi-feature data.

2.4. Extreme Value Theory

Extreme Value Theory (EVT) is an important branch of probability theory and statistics, specializing in the statistical study of extreme value distributions at the tail of random variables. In fields such as finance, insurance, engineering and environmental science, extreme data often have significant impacts, making accurate statistical inference on them crucial. EVT enables statistical inference and modeling of the tail of a probability distribution using only tail characteristics of the data, without knowing the overall population distribution. This makes EVT highly valuable for modeling and predicting rare extreme data.

In the wind power prediction problem addressed in this paper, conventional prediction models suffer from extremely large errors in two types of extreme scenarios: 1. sudden power spikes; 2. abrupt power drops. These extreme data account for a very small proportion of the overall dataset and belong to the distribution tail, which cannot be accurately characterized by ordinary least squares. EVT specializes in modeling tail extreme data, can quantify the boundary of extreme prediction errors, produce more reasonable prediction intervals, and support power grid dispatching security.

The Peaks Over Threshold (POT) method is adopted to model only the tail with the largest prediction errors, paired with the Generalized Pareto Distribution (GPD) to fit the distribution pattern of extreme errors, and ultimately determine the prediction interval covering extreme risks. The algorithm flow is as follows:

(1) Calculate residuals: Compute absolute residuals based on the point prediction results of the validation set.

(2) Determine the threshold: Take the 90th percentile of the validation set absolute residuals as the threshold u , and retain only error samples exceeding the threshold.

(3) Distribution fitting: Fit the Generalized Pareto Distribution GPD (ξ, σ) to the excess values above the threshold, where ξ is the shape parameter and σ is the scale parameter, estimated via maximum likelihood estimation.

(4) Construct prediction interval: Derive the extreme error quantile q_{EVT} at confidence level $1 - \alpha$ from the fitted GPD. The prediction interval is then $[\hat{P}_{t+h} - q_{EVT}, \hat{P}_{t+h} + q_{EVT}]$.

3. Experimental design

3.1. Data preprocessing

This paper conducts experiments using 10-minute resolution GridScientific operating data from a public wind farm covering January to April 2026. The dataset includes wind farm-level grid-side statistical variables such as active power, reactive power, voltage and current. ActivePowerMean is taken as the prediction target variable.

The data preprocessing steps are as follows: (1) Merge CSV files from January to April 2026; (2) Sort by timestamp and remove duplicate time records; (3) Eliminate or set to null abnormal status records where WPSStatus $\neq 0$; (4) Truncate negative active power values to 0; (5) Resample the data into a regular 10-minute time series; (6) Perform linear interpolation on missing values and fill endpoints with neighboring values; (7) Standardize input features using statistics from the training set. The dataset is divided chronologically into training set, validation set and test set with a ratio of 70%:15%:15% to avoid future information leakage in time series prediction.

3.2. Model flow

The overall workflow of the model proposed in this paper is illustrated in Figure 1, and the specific steps are as follows.

Step 1: Preprocessing. Preprocess raw wind power and supporting electrical quantity time series data.

Step 2: Prediction sample construction. Construct training and prediction samples, using data from the past 4 hours as input to predict power for the next 10, 30, and 60 minutes.

Step 3: Framework construction. Build the persistence model baseline branch and the difference ridge regression backbone branch. The backbone branch includes three methods: single-feature ridge regression, multi-feature ridge regression and difference ridge regression; the most suitable method is selected for modeling.

Step 4: Point prediction fusion. Automatically adjust model weights based on historical validation performance, and combine multiple prediction results via weighting to obtain more stable point prediction values.

Step 5: Extreme error modeling. Apply extreme value theory to analyze extreme scenarios with large errors and fit the fluctuation pattern of extreme errors.

Step 6: Prediction interval construction. Superimpose the extreme error range on point predictions to generate 95% confidence power prediction intervals.

Step 7: Output final results. Output both wind power point prediction values and probabilistic prediction intervals as the final wind power prediction output.

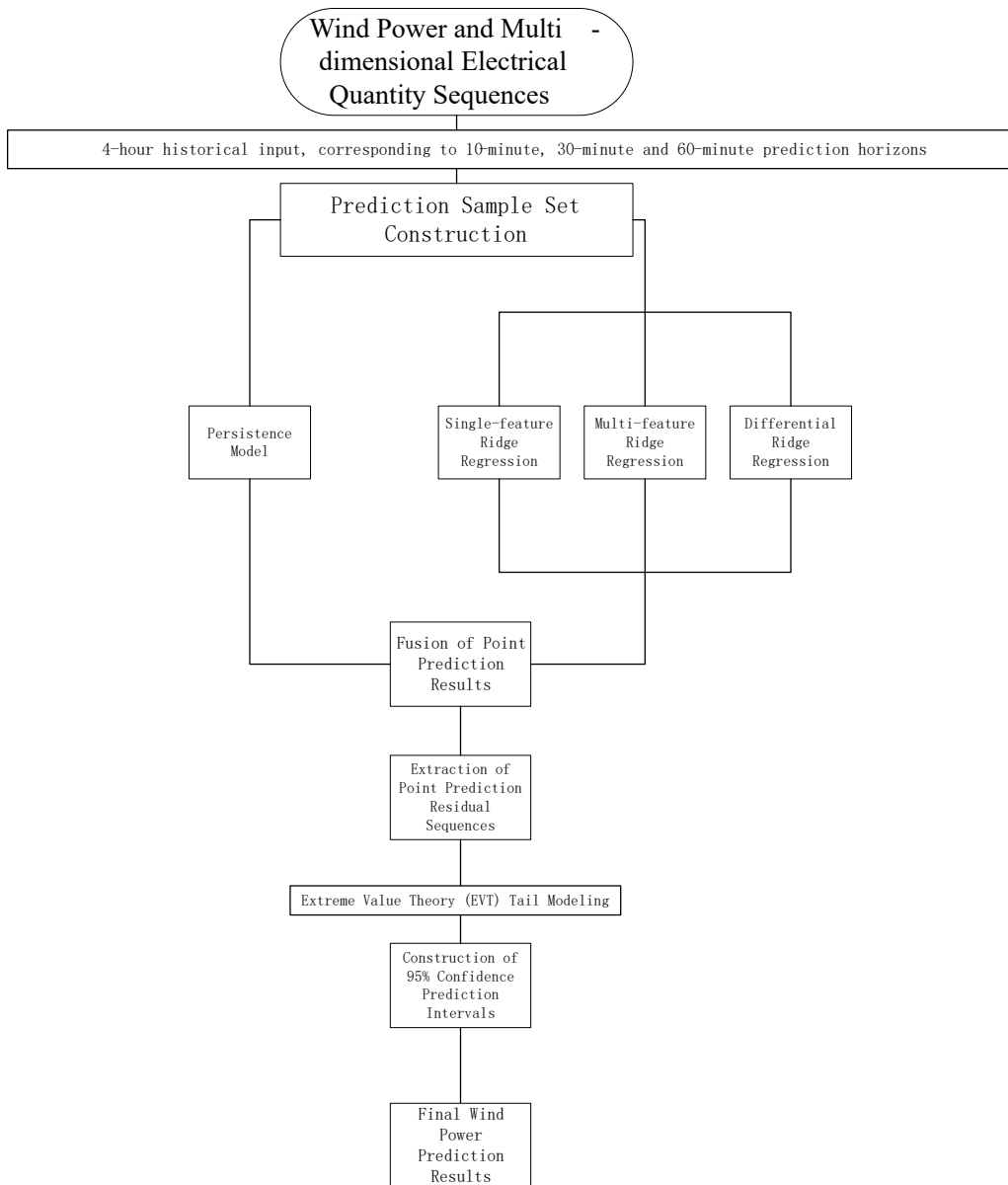


Figure 1. Model prediction flowchart

3.3. Comparison models

Table 1. Settings of comparison models

Model	Description
Persistence	Persistence prediction model
PowerOnly Ridge-AR	Ridge regression autoregressive model using only historical active power
MultiFeature Ridge	Direct ridge regression model using multi-feature historical windows
MultiFeature Delta-Ridge	Ridge regression model using multi-feature historical windows to predict power differences

Table 1. (continued)

Proposed-MFBlend-EVT	The multi-feature adaptive fusion and EVT tail calibration model proposed in this paper
----------------------	---

As shown in Table 1, the first four models are mainly used for point prediction comparison, while Proposed-MFBlend-EVT outputs both point prediction results and EVT-calibrated 95% prediction intervals.

4. Experimental results and analysis

4.1. Data description

This paper takes wind power operation data from a public wind farm as the research object, with a time resolution of 10 minutes. Wind farm-level grid-side data from January to April 2026 is selected as the original dataset, including multiple electrical indicators such as active power, reactive power, voltage and current. The time series plot of active power from part of the raw data is shown in Figure 2. After data processing and sample construction, the multi-feature ridge regression fusion and extreme value theory calibration prediction model is built, and multiple comparative experiments are conducted to verify the effectiveness of the proposed method in short-term wind power prediction.

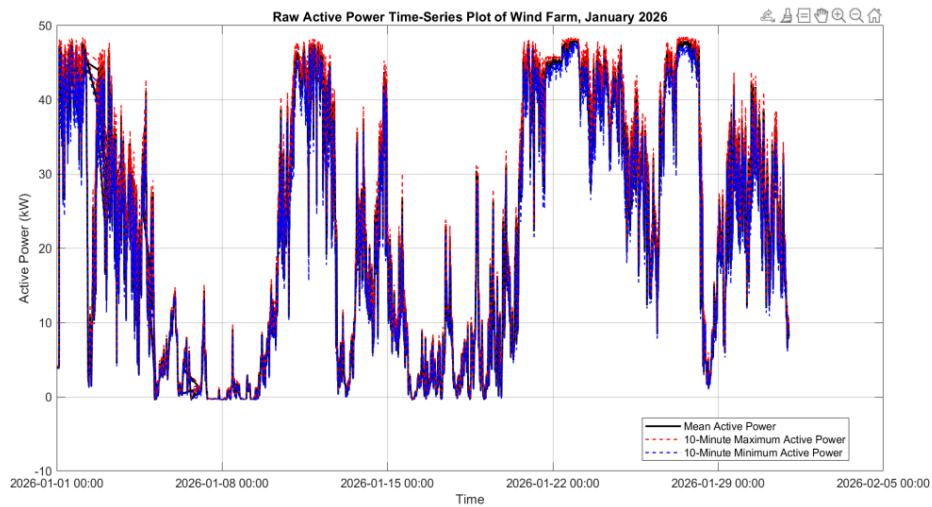


Figure 2. Raw data

4.2. Adaptive model selection results by prediction horizon

Table 2. Fusion model selection results under different prediction horizons

Prediction Horizon	Selected Backbone Model	Persistence Model Weight α	Backbone Model Weight $1-\alpha$	EVT Method
10 min	MultiFeature Delta-Ridge	0.30	0.70	GPD

Table 2. (continued)

30 min	MultiFeature Delta-Ridge	0.09	0.91	GPD
60 min	MultiFeature Ridge	0.23	0.77	GPD

As shown in Table 2, the selected backbone models and fusion weights vary across prediction horizons. In 10-minute prediction, the persistence model retains a weight of 0.30, indicating obvious inertia in very short-term wind power. In 30-minute prediction, the weight of MultiFeature Delta-Ridge rises to 0.91, showing that as the prediction horizon increases, relying only on current power becomes insufficient and more multi-feature difference corrections are needed. In 60-minute prediction, MultiFeature Ridge is selected as the backbone, indicating that for longer horizons, directly predicting future power using multi-dimensional historical features is more stable.

4.3. Point prediction result analysis

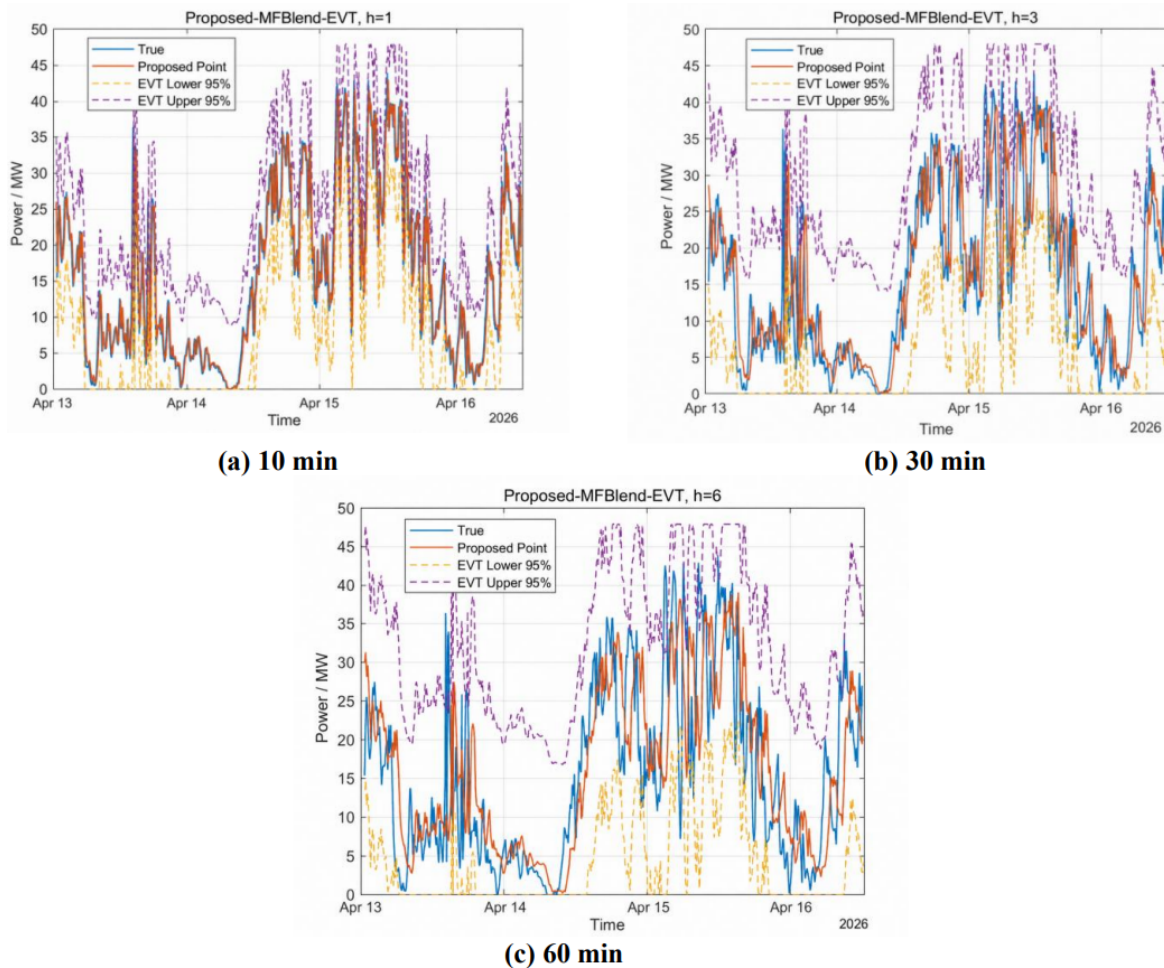


Figure 3. Point prediction results and EVT-calibrated 95% prediction intervals of the MFBblend-EVT model under different prediction horizons

As shown in Table 3 and Figure 3, In terms of MAE, Proposed-MFBblend-EVT achieves the best performance across all three prediction horizons, with MAE values of 1.6283 MW, 2.6740 MW and

3.4751 MW for 10-minute, 30-minute and 60-minute predictions respectively. Compared with the persistence model, the proposed method reduces MAE by approximately 1.35% and 2.06% in 30-minute and 60-minute prediction tasks, demonstrating that the multi-feature fusion strategy can improve mean absolute prediction error.

In terms of RMSE, although the proposed method is not optimal for every prediction horizon, it is very close to the best single-model result. Specifically, the optimal RMSE for 10-minute prediction is 3.0176 MW versus 3.0256 MW for the proposed method; for 30-minute prediction, the optimal RMSE is 4.5879 MW versus 4.5945 MW for the proposed method; for 60-minute prediction, the optimal RMSE is 5.6478 MW versus 5.6741 MW for the proposed method. Across all three horizons, the relative gap between the proposed method and the optimal RMSE is below 0.5%, indicating that while achieving the lowest MAE, the method still maintains favorable squared error performance.

Table 3. Comparison of point prediction errors across models

Prediction Horizon	Model	MAE/MW	RMSE/MW
10 min	Persistence	1.6286	3.0690
10 min	PowerOnly Ridge-AR	1.7167	3.0250
10 min	MultiFeature Ridge	1.6951	3.0352
10 min	MultiFeature Delta-Ridge	1.6407	3.0176
10 min	Proposed-MFBlend-EVT	1.6283	3.0256
30 min	Persistence	2.7106	4.7924
30 min	PowerOnly Ridge-AR	2.8904	4.5962
30 min	MultiFeature Ridge	2.7112	4.6055
30 min	MultiFeature Delta-Ridge	2.6818	4.5879
30 min	Proposed-MFBlend-EVT	2.6740	4.5945
60 min	Persistence	3.5482	6.1313
60 min	PowerOnly Ridge-AR	3.8569	5.7327
60 min	MultiFeature Ridge	3.5381	5.6478
60 min	MultiFeature Delta-Ridge	3.4980	5.7470
60 min	Proposed-MFBlend-EVT	3.4751	5.6741

4.4. Tail error and ramp sample result analysis

Table 4. Comparison of tail error and ramp sample errors

Prediction Horizon	Model	Tail-RMSE/MW	Ramp-MAE/MW
10 min	Persistence	8.6661	10.3227
10 min	MultiFeature Ridge	8.4385	9.8911
10 min	MultiFeature Delta-Ridge	8.4483	9.9949

Table 4. (continued)

10 min	Proposed-MFBlend-EVT	8.5006	10.0932
30 min	Persistence	13.1048	11.7645
30 min	PowerOnly Ridge-AR	12.3632	10.7036
30 min	MultiFeature Ridge	12.4389	10.6967
30 min	Proposed-MFBlend-EVT	12.4596	10.9793
60 min	Persistence	16.6961	12.9861
60 min	MultiFeature Ridge	14.6844	10.9667
60 min	Proposed-MFBlend-EVT	14.9895	11.4214

As shown in Table 4, Proposed-MFBlend-EVT does not consistently outperform all models in Tail-RMSE and Ramp-MAE, but it shows significant improvement over the persistence model. Particularly for the 60-minute horizon, the proposed method reduces Tail-RMSE from 16.6961 MW to 14.9895 MW and Ramp-MAE from 12.9861 MW to 11.4214 MW, indicating that the multi-feature fusion model can mitigate large errors and ramp sample prediction errors at longer prediction horizons.

It should be noted that the optimization objectives of Tail-RMSE and Ramp-MAE are not fully aligned with MAE. The fusion weights in this paper are selected primarily based on the comprehensive error objective of the validation set, so the final model leans toward overall average prediction accuracy rather than separately minimizing tail sample errors or ramp sample errors. This also reflects a certain trade-off between average error optimization and extreme sample optimization in wind power prediction.

4.5. EVT prediction interval result analysis

Table 5. Results of EVT-calibrated prediction intervals

Prediction Horizon	PICP_EVT95	PINAW_EVT95	EVT Radius
10 min	0.9780	0.3720	0.1860
30 min	0.9768	0.5814	0.2907
60 min	0.9784	0.6895	0.3448

As shown in Table 5, the EVT-calibrated prediction intervals achieve high coverage rates across all three horizons, with PICP values of 0.9780, 0.9768 and 0.9784 respectively — all above the target confidence level of 0.95. This indicates that GPD fitting based on validation set residual tails can effectively cover large error risks in the test set.

Meanwhile, PINAW increases with the prediction horizon, rising from 0.3720 for 10-minute prediction to 0.6895 for 60-minute prediction. This pattern conforms to the practical law of wind power prediction: the longer the prediction horizon, the greater the uncertainty of future wind conditions and power changes, so a wider prediction interval is required to cover potential errors. Although the EVT interval coverage is slightly higher than 0.95 — indicating a degree of conservatism — high-coverage intervals have practical value given the high risk that extreme wind power fluctuations pose to grid dispatching.

5. Conclusion

Targeting short-term wind power prediction, this paper adopts multi-feature difference ridge regression to obtain optimal wind power point prediction results, and applies extreme value theory to analyze tail errors and construct prediction intervals. Prediction verification for three horizons (10 min, 30 min, 60 min) is completed based on public wind farm data. The contributions of this paper are summarized as follows:

- (1) A multi-feature historical window representation method for short-term wind power prediction is constructed, incorporating grid-side statistical variables such as active power, reactive power, voltage and current into prediction inputs;
- (2) A prediction-horizon adaptive fusion strategy is proposed, which automatically determines the fusion weights of the persistence model and candidate ridge regression models on the validation set, improving the stability of point prediction across different horizons;
- (3) Extreme value theory is introduced to calibrate prediction residual tails, and 95% prediction intervals are constructed to characterize tail uncertainty in wind power prediction;
- (4) Experimental verification across 10 min, 30 min and 60 min horizons shows that the proposed model achieves optimal MAE performance and maintains high interval coverage.

References

- [1] Min, F., Ma, H., & Wang, Y. (2019). Dynamic sliding mode control for chaotic oscillation in power system with power disturbance. *Journal on Communications*, 40(1), 119-129.
- [2] Sun, W., & Wang, Y. W. (2018). Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network. *Energy Conversion and Management*, 157, 1-12.
- [3] Liu, H., & Chen, W. (2026). Short-term wind power prediction based on secondary decomposition cooperative TCNN-GRU. *Journal of Henan Polytechnic University (Natural Science Edition)*, 45(5), 11-19. DOI: 10.16186/j.cnki.1673-9787.2025100053.
- [4] Zheng, Z., Zhang, W., & Yan, R. (2026). Short-term wind power prediction based on ISOA-SVM. *Journal of Shenyang Institute of Engineering (Natural Science Edition)*, 22(3), 67-74+96. DOI: 10.13888/j.cnki.jsie(ns).2026.03.009.
- [5] Liu, X., Wang, Z., Wang, S., et al. (2026). Short-term wind power prediction based on adaptive decomposition and frequency domain enhanced iTransformer. *Power System Protection and Control*, 54(13), 103-113. DOI: 10.19783/j.cnki.pspc.260045.
- [6] Zhang, Y., Zhang, J., Ji, X., et al. (2026). Short-term wind power prediction for extreme weather based on secondary modal decomposition and DA-iTransformer. *Distribution & Utilization*, 43(6), 26-40. DOI: 10.19421/j.cnki.1006-6357.2026.06.003.
- [7] Xu, M., Chen, F., Hong, H., et al. (2026). Distributed photovoltaic power generation feature monitoring method based on SVDD and ARIMA fusion model. *Information Technology*(7), 169-174. DOI: 10.13274/j.cnki.hdzj.2026.07.029.
- [8] Wang, Q., Zhao, X., & Zhang, F. (2026). Research on change and prediction of acoustic environmental quality in Anhui Province based on ARIMA method. *Environmental Monitoring and Forewarning*, 18(4), 56-65.
- [9] Wang, G., Hou, H., Lan, J., et al. (2026). Glass performance prediction based on sample data multiple linear regression and quadratic summation model. *Glass*, 53(7), 13-17.
- [10] Wang, D., & Guo, W. (2026). Prediction of cargo throughput in Beibu Gulf Port based on grey GM-BP neural network combination model. *Journal of Nanning Normal University (Natural Science Edition)*, 1-11. <https://doi.org/10.16601/j.cnki.issn2096-7330.2026.05.008>.
- [11] Wang, Q., & Ruan, L. (2026). Research on groundwater quality evaluation model based on MBSO-BP neural network. *China Computer & Communication*, 38(16), 32-34.
- [12] Qiao, S., Xiong, J., & Chang, Y. (2026). The impact of climate physical risks on China's bulk commodity market — Analysis based on quantile and random forest. *Contemporary Economy*, 1-12. <https://link.cnki.net/urlid/42.1430.F.20260730.0933.010>.

- [13] Yang, Y., Yao, C., Li, S., et al. (2026). Spatiotemporal pattern and driving factors of drought and flood disasters in rice production in the Yangtze River Economic Belt: Analysis based on random forest-SHAP model. *Chinese Journal of Agricultural Resources and Regional Planning*, 1-17.
<https://link.cnki.net/urlid/11.3513.S.20260731.1025.002>
- [14] Wang, B. (2026). Research on multi-step prediction of offshore wind power resources based on GAT-LSTM and dynamic graph embedding. *Information Recording Materials*, 27(15), 89-91. <https://doi.org/10.16009/j.issn.1009-5624.2026.15.029>
- [15] Feng, K., Cao, S., Chen, Z., et al. (2026). Research on driver monitoring system algorithm based on convolutional neural network. *Automobile Applied Technology*, 51(14), 35-40+52. <https://doi.org/10.16638/j.cnki.1671-7988.2026.014.006>
- [16] He, Y., He, L., Fang, H., et al. (2026). Accuracy compensation of generalized decoupled parallel attitude adjustment platform based on ridge regression model. *Journal of Mechanical Engineering*, 1-13.
<https://link.cnki.net/urlid/11.2187.TH.20260630.1606.056>